Review of Bell et al.: "W-band Radar Observations for Fog Forecast Improvement: an Analysis of Model and Forward Operator Errors"

This paper presents an analysis of model and forward operator errors when assimilating 95 GHz reflectivity profiles to improve fog forecasts. This is a very interesting approach, which follows from earlier work from Borderies et al. work. The paper is very well written and structured. I do have two main comments, which I believe need to be addressed before the paper can be accepted, and they both relate to the same general issue. Then I only have a list of relatively minor comments.

Major comment:
My only major issue with the paper is related to how the assimilation of W-band reflectivities will actually look like in practice. The current version of the paper confused me in that respect, mostly sections 4.1 and 4.2. I was a bit confused first in section 4.1. What is the purpose of this exercise in the authors' mind ? I though initially that you wanted to demonstrate that the information content of reflectivity profiles to constrain assumptions in the three drop size distribution parameters was high, and you do show that the simulated reflectivity profiles do change substantially with the droplet concentration. In the retrieval world, such large change in the reflectivity profile as a function of a free parameter is exactly what you hope for. But you don't present things like that, you talk about this result as an "uncertainty". So here's my problem, which goes back to the main objective of this study, which is to demonstrate the potential impact of assimilating reflectivity. How is that going to work? What are going to be the increments produced by the reflectivity constraint? I thought you would free up some of the DSD parameters to adjust them to the observed reflectivity profile. But you don't seem to plan to do that, otherwise you would not treat this variability as an error of your forward model.
What I think this means is that the paper should include some description of how the assimilation is going to work.
I reached the same conclusion when reading section 4.2, explaining the MRP process. I am a bit worried by this approach, but maybe I shouldn't, as it depends how you are going to use that in practice. I thought initially that this was a process to shift the whole time series of model profiles by a given amount, so that you can take into account a possible mismatch in time with the initial development of the fog layer. But here you take any profile within a 6h window that matches the observations, which means that you completely lose the spatial continuity of the simulated fog layer in time, it can be any point of the 28*28km domain and it will change from one time step to another. How is that going to help in real situations when you assimilate Z profiles?
Again, in order to put your very interesting results in perspective of how you are going to use them when assimilating the reflectivity profiles for real, you need a section or a paragraph explaining how it's going to work.

Minor comments:

1. Lines 29-30: you say that this has been increasingly discussed "in recent years" but you cite a paper from 11 years ago, so it reads a bit funny. Any more recent discussions to justify your claim?
2. Line 41 : what is "advance time" ? I suspect you talk about "lead time" here ?
3. Line 42: I suggest "developed" instead of "seen"
4. Line 55-56 : what lead time of simulation are you using for this ?
5. Lines 60-61: why is the forward approach better posed than the Z - LWC conversion ("backward approach"_ ? The problem with Z to LWC is when there are few drizzle-sized drops in the volume, but I assume that in fog layers this is not so much of an issue, so the

Z to LWC approach trained with observations may be as accurate as the other way around. Please clarify why you think that's the case.

6. Title of section 2.2 : should BASTA be in capital letters ?
7. Line 102 : Fourth (not forth)
8. Lines 108-109: it's a bit inaccurate, I would rather say : "the fact that BASTA has separate transmitting and receiving antennas …"
9. Paragraph starting line 200: I think it would be useful to remind the reader what is expected to be the perfect score for each index used and where you consider that skills are not satisfying.
10. I have a suggestion about this analysis of scores. Showing just scores seems a bit incomplete and missed opportunity to dig more into the model performance. For instance, I would have liked to see PDFs of visibility from the model when fog is observed. It could be that the model is very slightly underestimating or overestimating visibility, which could change the scores substantially when using a hard limit of max visibility of 1km. We need to know how well the model predicts visibility overall in observed fog conditions. That would also provide quantitative information to the modellers to improve parameterizations. Also, because you have different kinds of fog, would you have enough cases to provide those scores and visibility PDFs for different types of fog you introduced earlier?
11. Line 211-212: I don't believe that's what a CSI of 0.32 means. This is how to interpret POD, not CSI. This needs to be addressed.
12. Figure 1 and associated text: While showing an example (good idea), it would be interesting to add two panels with vertical cross-sections of observed and modelled fog at the SIRTA site as well. Later, once you have developed your matching technique you could show the result on the same vertical cross-section to demonstrate the improvement visually as well.
13. Line 237: " 12-hour window". Isn't that too big a window ? Atmospheric conditions can change substantially, as well as radiative forcing over such a large window. As a result, you could get fog in the model developing from very different processes.
14. Lines 247-248: it is very interesting to learn that the model tends to produce shorter fog episodes. Is that only when fog episodes are both observed and simulated or separately for all observed and all simulated ? The reason why I ask is because you say that the model tends to produce a lot of cases that are not observed. What would that number be for the simulated when observed and simulated when not observed, any difference ?
15. Figure 3 and its interpretation. I think your statement that simulated fog top tends to be larger (and you should say higher not larger here) is not obvious from the figure. I'd rather stick to the bias value, maybe say that distribution of differences (is it model – obs by the way ?) is centred on zero with the std of 104m.
16. Paragraph starting line 336. Those are very interesting results, as is the impact on reflectivity from Fig.5. However, maybe some combination of the possible values of these three parameters is unrealistic. Does the literature say anything about the co-variability of these parameters? For instance are high concentrations always associated with a narrower range of nu or the other parameter ? That would help build a more realistic picture of the true variability of the reflectivity simulations if you can use such knowledge in your sensitivity analysis.
17. Line 377: " occur around 90 min". Is that your visual inspection estimate or is it the value that your MRP process found ?
18. Line 379: "Figure 7"

19. Figure 6: I think you should mask out the profiles with precipitation in both observations and model (using grey squares or something), as it is distracting from the main message of the figure.
20. Line 444 and Figure 10: One thing you should mention from this Figure, which is very positive, is that the distribution of errors is a lot narrower and around zero when MRP is used.

Good luck with the review,
Alain Protat
Melbourne, 29/01/2021